



2570  
0830

## **RAW SEQUENCE LISTING** **ERROR REPORT**

The Biotechnology Systems Branch of the Scientific and Technical Information Center (STIC) detected errors when processing the following computer readable form:

Application Serial Number: 10/029,988  
Source: OIPÉ  
Date Processed by STIC: 9/4/02

THE ATTACHED PRINTOUT EXPLAINS DETECTED ERRORS.

PLEASE FORWARD THIS INFORMATION TO THE APPLICANT BY EITHER:

- 1) INCLUDING A COPY OF THIS PRINTOUT IN YOUR NEXT COMMUNICATION TO THE APPLICANT, WITH A NOTICE TO COMPLY or,
- 2) TELEPHONING APPLICANT AND FAXING A COPY OF THIS PRINTOUT, WITH A NOTICE TO COMPLY

FOR CRF SUBMISSION QUESTIONS, PLEASE CONTACT MARK SPENCER, 703-308-4212.

FOR SEQUENCE RULES INTERPRETATION, PLEASE CONTACT ROBERT WAX, 703-308-4216.

PATENTIN 2.1 e-mail help: [patin21help@uspto.gov](mailto:patin21help@uspto.gov) or phone 703-306-4119 (R. Wax)

PATENTIN 3.0 e-mail help: [patin3help@uspto.gov](mailto:patin3help@uspto.gov) or phone 703-306-4119 (R. Wax)

TO REDUCE ERRORED SEQUENCE LISTINGS, PLEASE USE THE **CHECKER**  
**VERSION 3.1 PROGRAM**, ACCESSIBLE THROUGH THE U.S. PATENT AND  
TRADEMARK OFFICE WEBSITE. SEE BELOW FOR ADDRESS:

<http://www.uspto.gov/web/offices/pac/checker>

Applicants submitting genetic sequence information electronically on diskette or CD-Rom should be aware that there is a possibility that the disk/CD-Rom may have been affected by treatment given to all incoming mail.

Please consider using alternate methods of submission for the disk/CD-Rom or replacement disk/CD-Rom.

Any reply including a sequence listing in electronic form should NOT be sent to the 20231 zip code address for the United States Patent and Trademark Office, and instead should be sent via the following to the indicated addresses:

1. EFS-Bio (<<http://www.uspto.gov/ebs/efs/downloads/documents.htm>> , EFS Submission User Manual - ePAVE)
2. U.S. Postal Service: U.S. Patent and Trademark Office, Box Sequence, P.O. Box 2327, Arlington, VA 22202
3. Hand Carry directly to:  
U.S. Patent and Trademark Office, Technology Center 1600, Reception Area, 7<sup>th</sup> Floor, Examiner Name, Sequence Information, Crystal Mall One, 1911 South Clark Street, Arlington, VA 22202  
Or  
U.S. Patent and Trademark Office, Box Sequence, Customer Window, Lobby, Room 1B03, Crystal Plaza Two, 2011 South Clark Place, Arlington, VA 22202
4. Federal Express, United Parcel Service, or other delivery service to: U.S. Patent and Trademark Office, Box Sequence, Room 1B03-Mailroom, Crystal Plaza Two, 2011 South Clark Place, Arlington, VA 22202

# Raw Sequence Listing Error Summary

ERROR DETECTED	SUGGESTED CORRECTION	SERIAL NUMBER: 101029,988
ATTN: NEW RULES CASES: PLEASE DISREGARD ENGLISH "ALPHA" HEADERS, WHICH WERE INSERTED BY PTO SOFTWARE		
1 _____ Wrapped Nucleics Wrapped Aminos	The number/text at the end of each line "wrapped" down to the next line. This may occur if your file was retrieved in a word processor after creating it. Please adjust your right margin to .3; this will prevent "wrapping."	
2 _____ Invalid Line Length	The rules require that a line not exceed 72 characters in length. This includes white spaces.	
3 <input checked="" type="checkbox"/> Misaligned Amino Numbering	The numbering under each 5 <sup>th</sup> amino acid is misaligned. Do not use tab codes between numbers; use space characters, instead.	
4 _____ Non-ASCII	The submitted file was not saved in ASCII(DOS) text, as required by the Sequence Rules. Please ensure your subsequent submission is saved in ASCII text.	
5 _____ Variable Length	Sequence(s) _____ contain n's or Xaa's representing more than one residue. Per Sequence Rules, each n or Xaa can only represent a single residue. Please present the maximum number of each residue having variable length and indicate in the <220>-<223> section that some may be missing.	
6 _____ PatentIn 2.0 "bug"	A "bug" in PatentIn version 2.0 has caused the <220>-<223> section to be missing from amino acid sequences(s) _____. Normally, PatentIn would automatically generate this section from the previously coded nucleic acid sequence. Please manually copy the relevant <220>-<223> section to the subsequent amino acid sequence. This applies to the mandatory <220>-<223> sections for Artificial or Unknown sequences.	
7 _____ Skipped Sequences (OLD RULES)	Sequence(s) _____ missing. If intentional, please insert the following lines for each skipped sequence: (2) INFORMATION FOR SEQ ID NO:X: (insert SEQ ID NO where "X" is shown) (i) SEQUENCE CHARACTERISTICS: (Do not insert any subheadings under this heading) (xi) SEQUENCE DESCRIPTION:SEQ ID NO:X: (insert SEQ ID NO where "X" is shown) This sequence is intentionally skipped  Please also adjust the "(ii) NUMBER OF SEQUENCES:" response to include the skipped sequences.	
8 _____ Skipped Sequences (NEW RULES)	Sequence(s) _____ missing. If intentional, please insert the following lines for each skipped sequence. <210> sequence id number <400> sequence id number 000	
9 _____ Use of n's or Xaa's (NEW RULES)	Use of n's and/or Xaa's have been detected in the Sequence Listing. Per 1.823 of Sequence Rules, use of <220>-<223> is MANDATORY if n's or Xaa's are present. In <220> to <223> section, please explain location of n or Xaa; and which residue n or Xaa represents.	
10 _____ Invalid <213> Response	Per 1.823 of Sequence Rules, the only valid <213> responses are: Unknown, Artificial Sequence, or scientific name (Genus/species). <220>-<223> section is required when <213> response is Unknown or is Artificial Sequence	
11 _____ Use of <220>	Sequence(s) _____ missing the <220> "Feature" and associated numeric identifiers and responses. Use of <220> to <223> is MANDATORY if <213> "Organism" response is "Artificial Sequence" or "Unknown." Please explain source of genetic material in <220> to <223> section. (See "Federal Register," 06/01/1998, Vol. 63, No. 104, pp. 29631-32) (Sec. 1.823 of Sequence Rules)	
12 _____ PatentIn 2.0 "bug"	Please do not use "Copy to Disk" function of PatentIn version 2.0. This causes a corrupted file, resulting in missing mandatory numeric identifiers and responses (as indicated on raw sequence listing). Instead, please use "File Manager" or any other manual means to copy file to floppy disk.	
13 _____ Misuse of n	n can only be used to represent a single nucleotide in a nucleic acid sequence. N is not used to represent any value not specifically a nucleotide.	



Does Not Comply  
Corrected Diskette Needed

OIPE

## RAW SEQUENCE LISTING

PATENT APPLICATION: US/10/029,988

DATE: 09/04/2002

TIME: 16:12:39

Input Set : A:\Sequence Listing.txt

Output Set: N:\CRF4\09042002\J029988.raw

3 &lt;110&gt; APPLICANT: Bio-Technology General Corp

*Mandatory*  
 <120> through <170> missing!  
 Per § 1.823(b) of the new  
 rule

## ERRORED SEQUENCES

*See sample sequence listing  
 attached.*

```

3878 <210> SEQ ID NO: 204
3879 <211> LENGTH: 266
3880 <212> TYPE: PRT
3881 <213> ORGANISM: Homo sapiens
3883 <400> SEQUENCE: 204
3885 Met Lys Tyr Leu Leu Pro Thr Ala Ala Ala Gly Leu Leu Leu Ala
3886 1 5 10 15
3888 Ala Gln Pro Ala Met Ala Glu Val Gln Leu Val Glu Ser Gly Gly
3889 20 25 30
3891 Val Val Arg Pro Gly Gly Ser Leu Arg Leu Ser Cys Ala Ala Ser Gly
3892 35 40 45
3894 Phe Thr Phe Asp Asp Tyr Gly Met Ser Trp Val Arg Gln Ala Pro Gly
3895 50 55 60
3897 Lys Gly Leu Glu Trp Val Ser Gly Ile Asn Trp Asn Gly Gly Ser Thr
3898 65 70 75 80
3900 Gly Tyr Ala Asp Ser Val Lys Gly Arg Phe Thr Ile Ser Arg Asp Asn
3901 85 90 95
3903 Ala Lys Asn Ser Leu Tyr Leu Gln Met Asn Ser Leu Arg Ala Glu Asp
3904 100 105 110
3906 Thr Ala Val Tyr Tyr Cys Ala Arg Met Arg Ala Pro Val Ile Trp Gly
3907 115 120 125
3909 Gln Gly Thr Leu Val Thr Val Ser Arg Gly Gly Gly Ser Gly Gly
3910 130 135 140
3912 Gly Gly Ser Gly Gly Gly Ser Ser Glu Leu Thr Gln Asp Pro Ala
3913 145 150 155 160
3915 Val Ser Val Ala Leu Gly Gln Thr Val Arg Ile Thr Cys Gln Gly Asp
3916 165 170 175
3918 Ser Leu Arg Ser Tyr Tyr Ala Ser Trp Tyr Gln Gln Lys Pro Gly Gln
3919 180 185 190
3921 Ala Pro Val Leu Val Ile Tyr Gly Lys Asn Asn Arg Pro Ser Gly Ile
3922 195 200 205
3924 Pro Asp Arg Phe Ser Gly Ser Ser Ser Gly Asn Thr Ala Ser Leu Thr
3925 210 215 220
3927 Ile Thr Gly Ala Gln Ala Glu Asp Glu Ala Asp Tyr Tyr Cys Asn Ser
3928 225 230 235 240
3930 Arg Asp Ser Ser Gly Asn His Val Val Phe Gly Gly Gly Thr Lys Leu
3931 245 250 255
3933 Thr Val Leu Gly Ala Ala Ala Lys Ala Lys

```

## RAW SEQUENCE LISTING

PATENT APPLICATION: US/10/029,988

DATE: 09/04/2002

TIME: 16:12:39

Input Set : A:\Sequence Listing.txt

Output Set: N:\CRF4\09042002\J029988.raw

E--> 3934 260 265  
E--> 3938 85  
E--> 3941 1  
E--> 3942 NY01 393404\_1.DOC

-delete

amino numbering misaligned  
see item 3 on enclosure summary sheet

VERIFICATION SUMMARY  
PATENT APPLICATION: US/10/029,988DATE: 09/04/2002  
TIME: 16:12:40Input Set : A:\Sequence Listing.txt  
Output Set: N:\CRF4\09042002\J029988.raw

L:0 M:282 W: Numeric Field Identifier Missing, <120> is required.  
L:0 M:282 W: Numeric Field Identifier Missing, <160> is required.  
L:21 M:283 W: Missing Blank Line separator, <400> field identifier  
L:32 M:283 W: Missing Blank Line separator, <400> field identifier  
L:43 M:283 W: Missing Blank Line separator, <400> field identifier  
L:54 M:283 W: Missing Blank Line separator, <400> field identifier  
L:65 M:283 W: Missing Blank Line separator, <400> field identifier  
L:76 M:283 W: Missing Blank Line separator, <400> field identifier  
L:105 M:283 W: Missing Blank Line separator, <400> field identifier  
L:116 M:283 W: Missing Blank Line separator, <400> field identifier  
L:127 M:283 W: Missing Blank Line separator, <400> field identifier  
L:138 M:283 W: Missing Blank Line separator, <400> field identifier  
L:149 M:283 W: Missing Blank Line separator, <400> field identifier  
L:172 M:283 W: Missing Blank Line separator, <400> field identifier  
L:183 M:283 W: Missing Blank Line separator, <400> field identifier  
L:194 M:283 W: Missing Blank Line separator, <400> field identifier  
L:205 M:283 W: Missing Blank Line separator, <400> field identifier  
L:216 M:283 W: Missing Blank Line separator, <400> field identifier  
L:227 M:283 W: Missing Blank Line separator, <400> field identifier  
L:250 M:283 W: Missing Blank Line separator, <400> field identifier  
L:261 M:283 W: Missing Blank Line separator, <400> field identifier  
L:272 M:283 W: Missing Blank Line separator, <400> field identifier  
L:283 M:283 W: Missing Blank Line separator, <400> field identifier  
L:294 M:283 W: Missing Blank Line separator, <400> field identifier  
L:356 M:283 W: Missing Blank Line separator, <400> field identifier  
L:451 M:283 W: Missing Blank Line separator, <400> field identifier  
L:504 M:283 W: Missing Blank Line separator, <400> field identifier  
L:515 M:283 W: Missing Blank Line separator, <400> field identifier  
L:556 M:283 W: Missing Blank Line separator, <400> field identifier  
L:645 M:283 W: Missing Blank Line separator, <400> field identifier  
L:674 M:283 W: Missing Blank Line separator, <400> field identifier  
L:705 M:257 W: Feature value mis-spelled or invalid, <221> Name/Key for SEQ ID#:36  
L:708 M:283 W: Missing Blank Line separator, <400> field identifier  
L:716 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:36 after pos.:32  
L:737 M:283 W: Missing Blank Line separator, <400> field identifier  
L:766 M:283 W: Missing Blank Line separator, <400> field identifier  
L:795 M:283 W: Missing Blank Line separator, <400> field identifier  
L:824 M:283 W: Missing Blank Line separator, <400> field identifier  
L:853 M:283 W: Missing Blank Line separator, <400> field identifier  
L:912 M:283 W: Missing Blank Line separator, <400> field identifier  
L:1030 M:283 W: Missing Blank Line separator, <400> field identifier  
L:1056 M:283 W: Missing Blank Line separator, <400> field identifier  
L:1085 M:283 W: Missing Blank Line separator, <400> field identifier  
L:1114 M:283 W: Missing Blank Line separator, <400> field identifier  
L:1143 M:283 W: Missing Blank Line separator, <400> field identifier  
L:1199 M:283 W: Missing Blank Line separator, <400> field identifier  
L:3453 M:283 W: Missing Blank Line separator, <400> field identifier  
L:3464 M:283 W: Missing Blank Line separator, <400> field identifier

VERIFICATION SUMMARY  
PATENT APPLICATION: US/10/029,988

DATE: 09/04/2002  
TIME: 16:12:40

Input Set : A:\Sequence Listing.txt  
Output Set: N:\CRF4\09042002\J029988.raw

L:3475 M:283 W: Missing Blank Line separator, <400> field identifier  
L:3486 M:283 W: Missing Blank Line separator, <400> field identifier  
L:3497 M:283 W: Missing Blank Line separator, <400> field identifier  
L:3508 M:283 W: Missing Blank Line separator, <400> field identifier  
L:3519 M:283 W: Missing Blank Line separator, <400> field identifier  
L:3530 M:283 W: Missing Blank Line separator, <400> field identifier  
L:3934 M:332 E: (32) Invalid/Missing Amino Acid Numbering, SEQ ID:204  
M:332 Repeated in SeqNo=204  
L:3942 M:333 E: Wrong sequence grouping, Amino acids not in groups!  
L:3942 M:330 E: (2) Invalid Amino Acid Designator, NUMBER OF INVALID KEYS:1  
L:3942 M:252 E: No. of Seq. differs, <211> LENGTH:Input:266 Found:267 SEQ:204  
L:0 M:203 E: No. of Seq. differs, <160> Number Of Sequences:Input (0) Counted (204)

# Sample sequence listing

<110> Smith, John; Smithgene Inc.

<120> Example of a Sequence Listing

<130> 01-00001

<140> PCT/EP98/00001  
<141> 1998-12-31

<150> US 08/999,999  
<151> 1997-10-15

<160> 6

<170> PatentIn version 2.0

<210> 1  
<211> 389  
<212> DNA  
<213> Paramecium sp.

<220>  
<221> CDS  
<222> (279)...(389)

<300>  
<301> Doe, Richard  
<302> Isolation and Characterization of a Gene Encoding a  
Protease from Paramecium sp.  
<303> Journal of Genes  
<304> 1  
<305> 4  
<306> 1-7  
<307> 1988-06-31  
<308> 123456  
<309> 1988-06-31

<400> 1  
agctgtagtc attcctgtgt cctcttctct ctgggcttct caccctgcta atcagatctc 60  
agggagagtg tcttgacct cctctgcctt tgcagcttca caggcaggca ggcaggcagc E 120  
tgatgtggca atgctggca gtgccacagg cttctcagcc aggccttaggg tgggttccgc 180  
cgcggcgcgg cggccctct cgcgtctctc tcgcgctct ctctcgctct cctctcgctc 240

## Appendix 3, page 2.

ggacctgatt aggtgagcag gaggagggggg cagtttagc atg gtt tca atg ttc agc 296  
Met Val Ser Met Phe Ser

ttg tct ttc aaa tgg cct gga ttc tgt ttg ttc gtt tgt ttg ttc caa 344  
Leu Ser Phe Lys Trp Pro Gly Phe Cys Leu Phe Val Cys Leu Phe Gln

tgt ccc aaa gtc ctc ccc tgt cac tca tca ctg cag ccg aat ctt 389  
Cys Pro Lys Val Leu Pro Cys His Ser Ser Leu Gln Pro Asn Leu

<210> 2  
<211> 37  
<212> PRT  
<213> Paramecium sp.

<400> 2  
Met Val Ser Met Phe Ser Leu Ser Phe Lys Trp Pro Gly Phe Cys Leu  
1 5 10 15  
Phe Val Cys Leu Phe Gln Cys Pro Lys Val Leu Pro Cys His Ser Ser  
20 25 30

Leu Gln Pro Asn Leu  
35

<210> 3  
<211> 11  
<212> PRT  
<213> Artificial Sequence

<220>  
<221> Designed peptide based on size and polarity to act as a linker between the alpha and beta chains of Protein XYZ.

<400> 3  
Met Val Asn Leu Glu Pro Met His Thr Glu Ile  
1 5 10

<210> 4  
<400> 4  
000

[Annex VIII follows]



identifiers and their accompanying information as shown in the following table. The numeric identifier shall be used only in the "Sequence Listing." The order and presentation of the items of information in the "Sequence Listing" shall conform to the arrangement given below. Each item of information shall begin on a new line and shall begin with the numeric identifier enclosed in angle brackets as shown. The submission of those items of information designated with an "M" is mandatory. The submission of those items of information designated with an "O" is optional. Numeric identifiers <110> through <170> shall only be set forth at the beginning of the "Sequence Listing." The following table illustrates the numeric identifiers.

Numeric Identifier	Definition	Comments and Format	Mandatory (M) or Optional (O)
<110>	Applicant	Preferably max. of 10 names; one name per line; preferable format: Surname, Other Names and/or Initials	M
<120>	Title of Invention		M
<130>	File Reference	Personal file reference	M, when filed prior to assignment of appl. number
<140>	Current Application Number	Specify as: US 07/999,999 or PCT/US96/99999	M, if available
<141>	Current Filing Date	Specify as: yyyy-mm-dd	M, if available
<150>	Prior Application Number	Specify as: US 07/999,999 or PCT/US96/99999	M, if applicable include priority documents under 35 USC 119 and 120
<151>	Prior Application Filing Date	Specify as: yyyy-mm-dd	M, if applicable
<160>	Number of SEQ ID NOs	Count includes total number of SEQ ID NOs	M
<170>	Software	Name of software used to create the Sequence Listing	O
<210>	SEQ ID NO: #:	Response shall be an integer representing the SEQ ID NO shown	M
<211>	Length	Respond with an integer M expressing the number of bases or amino acid residues	

<212>	Type	<p>Whether presented sequence molecule is DNA, RNA, or PRT (protein). If a nucleotide sequence contains both DNA and RNA fragments, the type shall be "DNA." In addition, the combined DNA/RNA molecule shall be further described in the &lt;220&gt; to &lt;223&gt; feature section.</p>	M
<213>	Organism	<p>Scientific name, i.e. Genus/species, Unknown or Artificial Sequence. In addition, the "Unknown" or "Artificial Sequence" organisms shall be further described in the &lt;220&gt; to &lt;223&gt; feature section.</p>	M
<220>	Feature	<p>Leave blank after &lt;220&gt;. &lt;221-223&gt; provide for a description of points of biological significance in the sequence.</p>	<p>M, under the following conditions: if "n," "Xaa," or a modified or unusual L-amino acid or modified base was used in a sequence; if ORGANISM is "Artificial Sequence" or "Unknown"; if molecule is combined DNA/RNA.</p>
<221>	Name/Key	<p>Provide appropriate identifier for feature, preferably from WIPO Standard ST.25 (1998), Appendix 2, Tables 5 and 6</p>	<p>M, under the following conditions: if "n," "Xaa," or a modified or unusual L-amino acid or modified base was used in a sequence</p>
<222>	Location	<p>Specify location within sequence; where appropriate state number of first and last bases/amino acids</p>	<p>M, under the following conditions: if "n," "Xaa," or a modified or unusual L-amino acid or modified</p>

<223>

Other Information

Other relevant information;  
four lines maximum

base was used in  
sequence

M, under the following conditions:  
if "n," "Xaa," or  
a modified or unusual L-amino acid  
r, modified base  
was used in a  
sequence; if  
ORGANISM  
is "Artificial  
Sequence" or  
"Unknown"; if  
molecule is combined DNA/RNA

<300>

Publication Information

Leave blank  
after <300>

0

<301>

Authors

Preferably max  
of ten named  
authors of publication; specify  
one name per line;  
preferable format:  
Surname, Other  
Names and/or  
Initials

0

<302>

Title

0

<303>

Journal

0

<304>

Volume

0

<305>

Issue

0

<306>

Pages

0

<307>

Date

Journal date on which  
data published;  
specify as yyyy-mm-  
dd, MM-yyyy or  
Season-yyyy

0

<308>

Database  
Accession  
Number

Accession number  
assigned by database  
including  
database name

0

<309>

Database Entry  
Date

Date of entry in  
database; specify  
as yyyy-mm-dd or  
MM-yyyy

0

<310>

Patent Document  
Number

Document number;  
for patent-type  
citations only.  
Specify as, for  
example, US  
07/999,999

0

1/29/99 1:51 PM

<311>	Patent Fil Date	Document filing date, for patent- type citations only; specify as yyyy-mm-dd	0
<312>	Publication Date	Document publication date, for patent-type citations only; specify as yyyy-mm-dd.	0
<313>	Relevant Residues	FROM (position) TO (position)	0
<400>	Sequence	SEQ ID NO should follow the numeric identifier and should appear on the line pre- ceding the actual sequence	M

5. Section 1.024 is revised to read as follows:

1.024 Form and format for nucleotide and/or amino acid sequence submissions in computer readable form.

(a) The computer readable form required by 1.021(c) shall meet the following specifications:

(1) The computer readable form shall contain a single "Sequence Listing" as either a diskette, series of diskettes, or other permissible media outlined in paragraph (c) of this section.

(2) The "Sequence Listing" in paragraph (a) (1) of this section shall be submitted in American Standard Code for Information Interchange (ASCII) text. No other formats shall be allowed.

(3) The computer readable form may be created by any means, such as word processors, nucleotide/amino acid sequence editors or other custom computer programs; however, it shall conform to all specifications detailed in this section.

(4) File compression is acceptable when using diskette media, so long as the compressed file is in a self-extracting format that will decompress on one of the systems described in paragraph (b) of this section.

(5) Page numbering shall not appear within the computer readable form version of the "Sequence Listing" file.

(6) All computer readable forms shall have a label permanently affixed thereto on which has been hand-printed or typed: the name of the applicant, the title of the invention, the date on which the data were recorded on the computer readable form, the operating system used, a reference number, and an application serial number and filing date, if known.

(b) Computer readable form submissions must meet these format requirements:

(1) Computer: IBM PC/XT/AT, or compatibles, or Apple Macintosh;

(2) Operating System: MS-DOS, Unix or Macintosh;